18.6 A Sub-200mV 6T SRAM in 0.13µm CMOS

Bo Zhai, David Blaauw, Dennis Sylvester, Scott Hanson

University of Michigan, Ann Arbor, MI

Subthreshold operation holds significant promise of ultra low energy operation for emerging applications such as environmental and biomedical sensing and supply-chain management. A key obstacle in subthreshold design is the lack of robust and dense ultra-low voltage memories. To address this problem, we present a 0.13µm 2kb 6T SRAM fully functional from 1.2V to 193mV. To our knowledge, this is the first 6T SRAM capable of operating substantially below threshold voltage (V_t).

Previous work investigating subthreshold SRAM design is presented in [2,5,6]. However, the implementations in [2] and [6] (which we will refer to as a mux-based memory) uses a tree-like readout path implemented with muxes and incurs significant area, power and performance overheads. The 10T SRAM presented in [5] operates down to 380mV in 65nm CMOS but requires a second power supply of 480mV for the write driver as well as memory cell redundancy.

The major difficulties with traditional 6T SRAM design in subthreshold operation include: 1) exponential dependence of drive current on V_t in subthreshold operation heightens the impact of process variations, 2) reduced I_{on}/I_{off} ratio in subthreshold operation compromises robustness (see Fig. 18.6.3) and 3) the sensing voltage of sense amplifiers does not scale with V_{DD}. Furthermore, the differential structure of traditional 6T SRAM tends to favor writability in subthreshold operation at the expense of readability, resulting in unbalanced failure rates. To address these issues, we propose a single-ended 6T SRAM with dynamic supply voltage suppression for improved writability. With no die-specific adjustments, the SRAM is fully functional at a supply voltage of 208mV while enabling on-chip tuning allows operation down to 193mV. Furthermore, the design uses equalized voltage suppression to reduce leakage while holding state.

Fig. 18.6.1 shows the architecture of the proposed 6T SRAM with 16 bitcells connected to one bitline. The single-ended structure has improved read stability while sacrificing write stability. To recover write stability, the circuit uses adjustable strength header and footer devices [7]. When a write occurs, only the narrower headers/footers are enabled by asserting *wr_en*, resulting in a temporary supply voltage droop which allows the stored state to be overwritten, as shown in Fig 18.6.2. To provide sufficient supply droop, a PMOS (NMOS) device is used as the footer (header). Despite the supply droop, the state of the non-accessed cells is retained, as illustrated by the simulated butterfly curves in Fig. 18.6.3.

If necessary, a footer controlled by *ft_bias* allows the droop on *virGND* and *virVDD* to be tuned individually for each die. Using an on-chip bias generator, *ft_bias* can be programmed to one of 6 values between 0 and V_{DD}. To address die-to-die V_t shifts, NMOS and PMOS body bias was also implemented.

In [4], it was shown that random dopant fluctuations (RDF) and die-to-die V_t shifts are dominant sources of drive strength variation in the subthreshold regime as opposed to gate length variation, due to reduced DIBL. Using Monte Carlo SPICE simulation, transistor sizes are determined as shown in Fig. 18.6.1 to meet robustness requirements, setting the minimum device width to 0.32µm. PMOS sizes are larger than in traditional 6T bit cell designs due to significantly increased NMOS/PMOS on-current ratio at low voltage, as shown in Fig 18.6.3. It was found that symmetrically sized feedback and forward inverters result in balancing read and write capability. Cell device sizes can be reduced if a less stringent supply voltage floor is needed or if memory design rules are available (only logic rules were used).

In the read-out path, a near-minimum sized inverter is used as the sensing element to reduce the bitline capacitance. The second inverter in the read-out path is larger for robustness and, based on simulation, is able to drive a tri-state line with up to 64 units. In the implemented design, the second level mux was restricted to 16 inputs, since 2kb is sufficient for the targeted sensor applications.

The timing generation is shown in Fig. 18.6.1 and is programmable to allow for improved performance. To address the increased variability in the subthreshold regime, we implemented both NAND-type and NOR-type pulse generation to achieve a tunable pulse window that extends beyond the half-cycle point.

Since the bitline is directly connected to the read-out inverter, static current can be high when the bitline is floating. Therefore *bl_charge_*, as well as *hold_en* and *wr_en* are asserted during standby mode. By enabling *hold_en*, the voltage droop in virGND and virVDD is increased and equalized, compared to the droop during a write access.

Fig. 18.6.4 shows frequency measurements for 4 chips fabricated in a 0.13µm bulk CMOS, showing the exponential dependence of frequency on V_{DD} in the subthreshold regime. The array operates at a frequency of 5.6MHz at 0.5V and 21.5kHz at 210mV. The measured energy per access for the proposed SRAM is compared with that of a mux-based memory [6] fabricated in the same technology with identical data and an activity rate of 0.5 access/cycle in both cases. For equal supply voltages, our SRAM consumes approximately 31% less energy with similar performance. The energy optimal supply voltage (V_{min}) lies at 340mV for our SRAM and at 400mV for the mux-based memory. The energy per access increases at supply voltages below V_{min} due to the exponential increase of circuit delay and the dominance of leakage current in this voltage regime. At the respective V_{min} voltages, our SRAM consumes 64% the energy of the mux-based memory. Furthermore, unlike the mux-based memory, V_{min} for the proposed SRAM matches more closely to that of a typical sensor processor core [6], thereby allowing both memory and core to operate at peak energy efficiency with a single supply voltage.

The area of the proposed 2kb SRAM is 28,600µm², which is nearly half that of the mux based memory at 54000µm². However, it is 42% larger than a commercial high-voltage SRAM, which fails below 720mV in our measurements. Fig 18.6.4 also shows the leakage measurement of the proposed SRAM both with and without asserting the *hold_en* signal, showing approximately a 2× leakage reduction.

Fig 18.6.5 shows the read, write and hold failure rates as a function of V_{DD} with and without the tuned footer (*ft_bias*) and/or body bias. Read and write failures are well balanced, as intended. Enabling the tuned footer does not significantly reduce the point of first failure, but does slow the increase in failure rate below this point, making redundancy more advantageous. Using both the tuned footer and body bias, the voltage of first failure is reduced to 193mV. Based on the measured failure rate, use of 2% redundancy can extend the minimum operating voltage to 170mV, to allow ultra low power operation. Fig 18.6.5 also shows that the first hold failure occurs at 136mV. Finally, Fig. 18.6.6 presents the energy versus frequency distribution of 14 tested dice for 3 operating voltages.

References:

[1] B. Zhai, D. Blaauw, D. Sylvester, et al., "Theoretical and Practical Limits of Dynamic Voltage Scaling," *DAC*, pp. 868-873, Jun., 2004.

[2] A. Wang and A. Chandrakasan, "A 180mV FFT Processor Using Subthreshold Circuit Techniques", *ISSCC Tech. Digest, Feb.,* pp. 292-293, Feb., 2004.

[3] L. Nazhandali, B. Zhai, A. Olson, et. al., "Energy Optimization of Subthreshold-Voltage Sensor Processors," *ISCA*, pp. 197-207, Jun., 2005.

[4] B. Zhai, S. Hanson, D. Blaauw, et al., "Analysis and Mitigation of Variability in Subthreshold Design", *ISLPED*, pp. 20-25, Aug., 2005.

[5] B. Calhoun and A. Chandrakasan, "A 256kb Sub-threshold SRAM in 65nm CMOS", *ISSCC Tech. Digest,* pp. 628-629, Feb., 2006.

[6] B. Zhai, et al., "A 2.60pJ/Inst Subthreshold Sensor Processor for Optimal Energy Efficiency", *Symp. VLSI Circuits*, pp. 154-155, Jun., 2006.

[7] K. Zhang, U. Bhattacharya, Z. Chen, et al., "SRAM Design on 65nm CMOS Technology with Integrated Leakage Reduction Scheme," Symp. *VLSI Circuits*, pp. 294-295, Jun., 2004.

 1-4244-0852-0/07/$25.00 ©2007 IEEE.

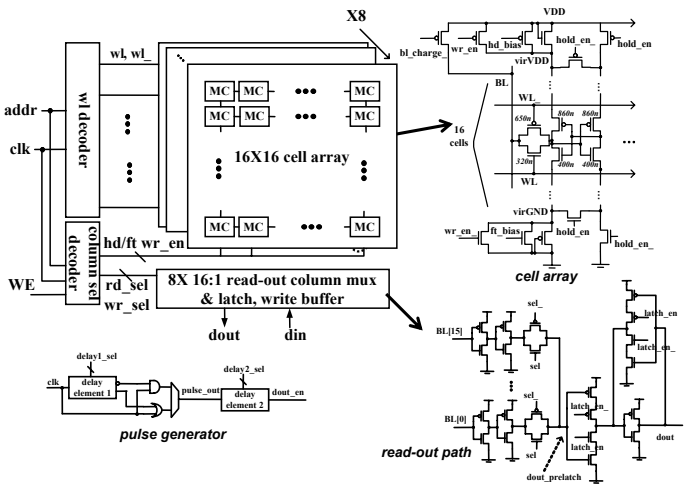


Figure 18.6.1: Top-level organization, read-out path and pulse generator. Read out path has a 16-to-1 muxthat is sized to ensure reliability. Pulse width and distribution delay are programmable to improve robustness to variability. 6T devicewidths are annotated.

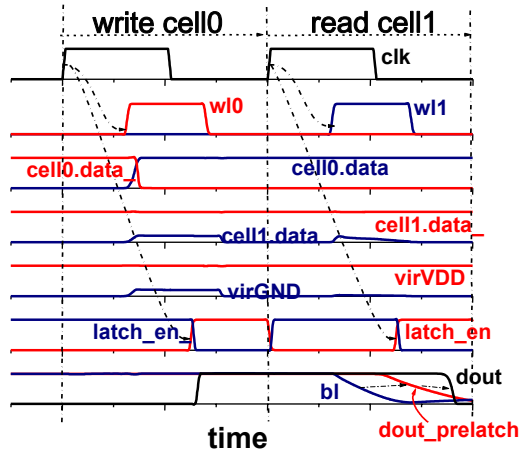


Figure 18.6.2: SPICE simulation of proposed 6T SRAM with a supplyvoltage of 250mV, showing consecutive read and write operation and supply voltage suppression through tunable headers and footers.

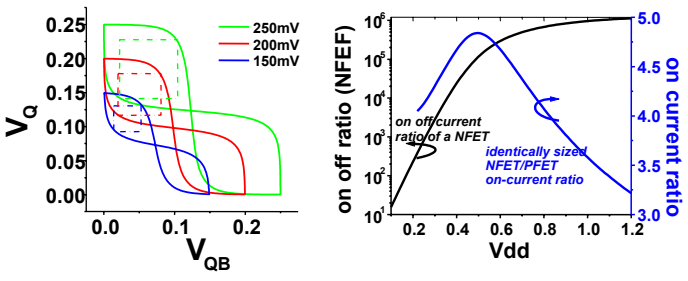


Figure 18.6.3: Butterfly curve simulation of proposed 6T cell at 250mV, 200mV, and 150mV power supply. On/off current ratio and NMOS/PMOS on-current ratio as a function of supply voltage .

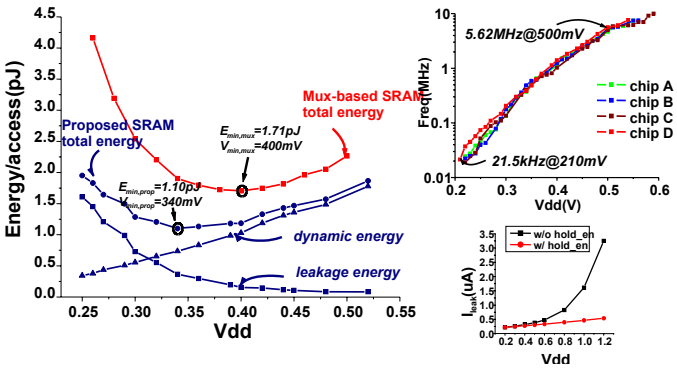


Figure 18.6.4: Frequency with V_{DD} scaling for 4 dice and energy consumption comparison to mux-based design [2] in the same technology. The proposed SRAM is about 1.6X more energy efficient than mux-based design with similar speed. Activity rate is 0.5 access/cycle.

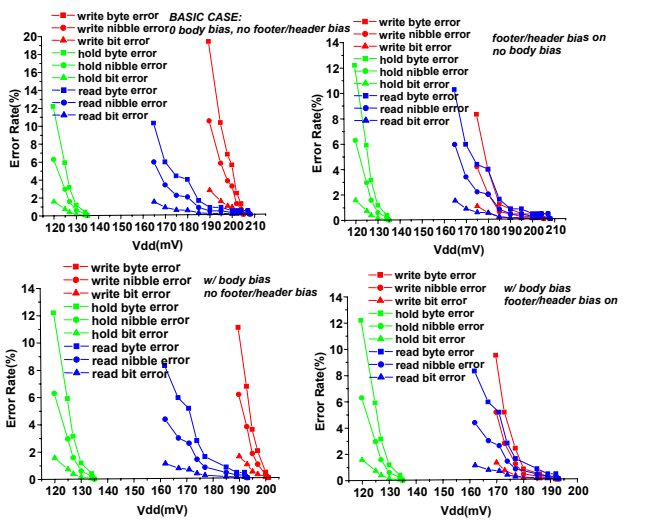


Figure 18.6.5: Read, write and hold failure rates with supply voltages when applying different techniques

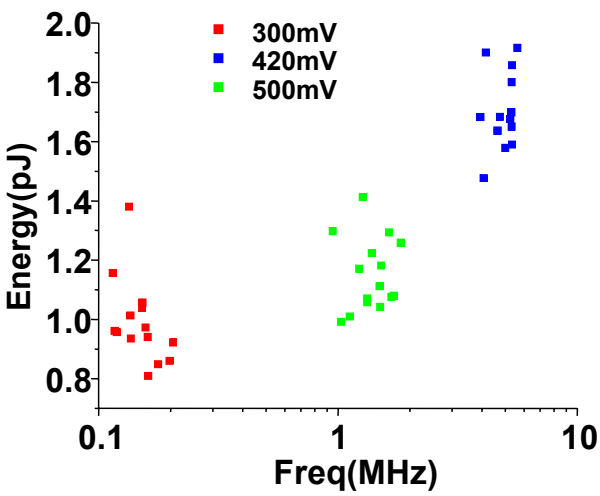
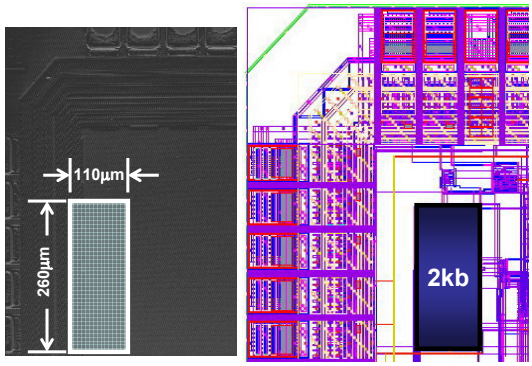


Figure 18.6.6 Frequency-energy scatter plot at different supply voltages. Activity rate is 1 access/cycle.

Continued on Page 606

18



Technology	0.13µm 8-metal CMOS
SRAM Size	2k bits
Area	28600µm2
Supply Voltage	1.2V-193mV
Frequency	5.62MHz@500mV 21.5kHz@210mV
Energy/Access (activity rate=0.5)	1.10pJ@340mV
Power	50nW@210mV

Figure 18.6.7: Die micrograph and layout.

 1-4244-0852-0/07/$25.00 ©2007 IEEE.